

MESCOLLEGE OF ENGINEERING, KUTTIAPPURAM
DEPARTMENT OF COMPUTER APPLICATIONS
20MCA245-MINI PROJECT

PROFORMA FOR THE APPROVAL OF THE THIRD SEMESTER MINI PROJECT

(Note: All entries of the proforma for approval should be filled up with appropriate and complete information.
Incomplete

Proforma of approval in any respect will be rejected.)

Mini Project Proposal No: ____1____
(Filled by the Department)

Academic Year : 2021-2022
Year of Admission: 2020

1. Title of the Project: ANALYSIS OF CT SCAN IMAGES TO PREDICT LUNG CANCER

STAGES USING IMAGE PROCESSING TECHNIQUES WITH SECURE PHR

2. Name of the Guide: MOHAMMAD JABIR

3. Number of the Student: MES20MCA-2037

4. Student Details (in BLOCK LETTERS)

Name: NSRINBP

Roll Number: 37

Signature 1.

Date: 1/11/2021

Approval Status: Approved/Not Approved

Signature of
Committee Members

Comments of The Mini Project Guide

Dated Signature

Initial Submission:

First Review :

Second Review :

Comments of The Project Coordinator

Dated Signature Initial Submission:

First Review

Second Review

Final Comments:

Dated Signature of

INTRODUCTION:

Lung cancer is one of the most dangerous and common cancer diseases in the world. Early detection of lung cancer can increase survival time of a patient. It is difficult for doctors to identify the cancer stages from Computed Tomography (CT) scan images. In this era of technology, a computer-aided system can help to predict lung cancer stages more accurately. Inspired by the recent success of image processing and machine learning techniques in the medical field, we have developed models using Gray level co-occurrence matrix (GLCM) based texture image analysis and Statistical parametric approach for helping doctors to detect lung cancer stages. Our approach involves image acquisition, preprocessing, feature extraction and finally classification. For feature extraction, two approaches are used: Gray level co-occurrence matrix (GLCM) based texture image analysis and Statistical parametric approach. For detecting lung cancer stages, four different classifiers are used and obtained the highest accuracy 78.95% with 0.77 precision and 0.83 recall using Support Vector Machine (SVM) in the Statistical parametric approach of feature selection.

Lung cancer is also familiar as lung carcinoma, caused by a malignant lung tumor which has uncontrolled cell growth. This cell growth can spread to the other part of the body by the metastasis process. Mainly, there are two types of lung cancer, one is small-cell lung cancer (SCLC) and the other one is non-small-cell lung cancer (NSCLC). The primary symptoms of lung cancer are coughing, losing weight, breath shortness and chest pain. One of the main reasons of lung cancer is smoking, besides being a passive smoker, air pollution and genetic factors are also responsible for lung cancer. Avoiding smoking with other risk factors can prevent lung cancer primarily. Lung cancer stage can be divided into limited stage and excessive stage. In limited stage, cancer is confined in one lung, and in excessive stage, cancer has spread to the other part of the body. CT images have been chosen as it is more efficient compared to X-ray for detecting lung cancer stages. The main reason to choose digital image processing techniques (DIP) for detecting lung cancer stage is that it gives better visualization and information compared to other forms. Image processing techniques are an easier way to analyze image cells and extract data from them. According to our knowledge, a few research works have been done on Lung Cancer stage detection. In this study, statistical parametric approach has the least amount of features compared to GLCM approach. As the number of features increases, the dimension increases, and then the power of number of features. Using statistical parametric approach, computation is not getting complex compared to GLCM approach. Statistical parametric approach gives the best result on our dataset. Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest and Naive Bayes classifiers are used for learning purposes. SVM with the highest accuracy 78.95% in Statistical parametric approach performs best for our selected dataset. The prime purpose of this research is to utilize the principles of data mining and data science in the domain of patient data. Due to the abundance and vast variety of general patient data, it is often overlooked. This research focuses on the complete life-cycle of medical health data from acquiring it to extracting valuable information from it. Through data mining using digital solutions, the process of collection of data becomes less crucial and with the advancement of data

storage technologies in terms of velocity and size make the process of data processing swift. The anonymity is yet another milestone in medical health data which has to be dealt with using techniques of cryptography and well protected data warehouses. Clinical data holds secrets of nation's healthcare. It can be used for comparisons and role models for the betterment of healthcare system. The adoption of smart digital solutions is being done to encourage individuals to proceed towards disease free world

Keywords Clinical Data, Cryptography, Data Collection, Health Data, Knowledge acquisition, Health solution

Existing system

Early detection of lung cancer is very important for successful treatment. And lung cancer can increase survival time of a patient. It is difficult for doctors to identify the cancer stages from Computed Tomography scan images. And also doctors need to ask their recent problems, medicines, what method of treatment they took

Proposed system

Right now we can actualize conclusion mining investigation utilizing Support Vector Machine calculation. Firstly, images were acquired then preprocessed. Preprocessing includes Smoothing, Enhancement, Segmentation, Morphological Opening and Selection of Region of interest (ROI). After that, features were extracted using GLCM based texture analysis and Statistical parametric approach then feature values were recalculated. Finally, we classified cancer stages as limited stage and extensive stage using classifiers and computed performance of classifiers. We can make doctors' job easier by using data science and deep learning technology. We collect all the details of the patients and display to doctors so they can easily understand the person's problem and give a better treatment. If a patient forgets to take any details, still a doctor can get it using the computer.

Basic functionalities:

Machine learning (ML)

ML is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

Datamining

Datamining is the process of finding anomalies, patterns and correlations within large datasets to mine information. Using a broad range of techniques, you can use this information to increase revenues, cut costs, improve customer relationships and reduce risks and more.

Cryptography

Cryptography is the study of secure communication techniques that allow only the sender and intended recipient of a message to view its contents. Here, data is encrypted using a secret key,

and then both the encoded message and secret key are sent to the recipient for decryption.

MODULE DESCRIPTION

- Admin
- User

Admin

- Login
- View Users
- View Feedback
- Add and manage data set
- Add and manage tips

USERS

- Registration
- Login
- Update profile
- View Tips
- Upload Image and View Result
- View Upload History
- Send Feedback

HARDWARE AND SOFTWARE REQUIREMENT Hardware Requirements

The selection of hardware is very important in the existence and proper working of any software. Then selection of hardware, the size and capacity requirements are also important.

Processor: Intel Pentium Core i3 and above, 64 bits

RAM: Min 3GB RAM

HARD DISK: 10GB

Software Requirements

One of the most difficult tasks is selecting software for the system, once the system requirements are found out then we have to determine whether a particular software package fits for those system requirements. The application requirement:

OPERATING SYSTEM: WINDOWS 10

FRONTEND: HTML, CSS, JAVASCRIPT

BACKEND: Mysql

IDEUSED:JetbrainsPycharm,Androidstudio

TECHNOLOGYUSED:PYTHONJAVA

FRAMEWORKUSED:Flask